Neurobiology of Disease

Cell Type-Specific Expression Analysis to Identify Putative Cellular Mechanisms for Neurogenetic Disorders

Xiaoxiao Xu,[1] Alan B. Wells,[2,3] David R. O'Brien,[2,3] Arye Nehorai,[1] and Joseph D. Dougherty[2,3]

[1]Preston M. Green Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, and Departments of [2]Genetics and [3]Psychiatry, Washington University School of Medicine in St. Louis, St. Louis, Missouri 63110

Recent advances have substantially increased the number of genes that are statistically associated with complex genetic disorders of the CNS such as autism and schizophrenia. It is now clear that there will likely be hundreds of distinct loci contributing to these disorders, underscoring a remarkable genetic heterogeneity. It is unclear whether this genetic heterogeneity indicates an equal heterogeneity of cellular mechanisms for these diseases. The commonality of symptoms across patients suggests there could be a functional convergence downstream of these loci upon a limited number of cell types or circuits that mediate the affected behaviors. One possible mechanism for this convergence would be the selective expression of at least a subset of these genes in the cell types that comprise these circuits. Using profiling data from mice and humans, we have developed and validated an approach, cell type-specific expression analysis, for identifying candidate cell populations likely to be disrupted across sets of patients with distinct genetic lesions. Using human genetics data and postmortem gene expression data, our approach can correctly identify the cell types for disorders of known cellular etiology, including narcolepsy and retinopathies. Applying this approach to autism, a disease where the cellular mechanism is unclear, indicates there may be multiple cellular routes to this disorder. Our approach may be useful for identifying common cellular mechanisms arising from distinct genetic lesions.

Introduction

The brain contains hundreds of distinct cell types, each with unique morphologies, projections, and functions. Yet, there are clear examples of neurological disruptions caused by deficiencies in just one cell type or circuit—such as dopaminergic neurons in Parkinson's disease, or hypocretinergic (Hcrt) neurons in narcolepsy. Clearly, distinct cell types in the nervous system contribute to different behaviors. However, the cellular disruptions that lead to the behavioral abnormalities in many disorders, including autism, are not clear. If there were a method to identify the cell types that serve as the intermediaries between a set of genetic lesions and a particular behavioral disruption, then one could identify cellular targets for treatment. Importantly, insights into the cell types responsible for a disorder create more obvious routes to treatment, as exemplified by the diverse strategies adopted for Parkinson's disease (Poewe et al., 2010; Rodriguez-Oroz, 2010; Pardal and López-Barneo, 2012).

How could variations in genes found in the genomes of every cell in the body lead to disruptions of only particular cell types? There are at least two potential explanations—selective vulnera-bility or selective expression (SE). Selective vulnerability posits that the physiological nature of a particular cell type makes it uniquely vulnerable to perturbations of a broadly expressed gene. SE posits that the gene expression of a cell type makes it uniquely disruptable—the variation disrupts a specifically expressed gene that is required only for the function of one particular cell type. At one extreme, a highly penetrant mutation in a gene expressed in only one cell type in the brain (e.g., a unique neurotransmitter; Peyron et al., 2000; Ercan-Sencicek et al., 2010) would strongly indicate disruption of that cell type as a mechanism for the disease.

There is a remarkable diversity of gene expression across cell types in the nervous system (Lein et al., 2007; Cahoy et al., 2008; Doyle et al., 2008; Dougherty et al., 2012), and there are clear examples from the literature of both SE and selective vulnerability. There are specific examples of disease genes from rare cases that are expressed only in one cell type, clearly implicating those cells and their circuits in a disorder (Peyron et al., 2000; Taheri et al., 2002; Ercan-Sencicek et al., 2010). And, there is the broader example of SE in the large number of mutations in rod- and cone-specific genes seen in genetic forms of blindness (Gal et al., 1994; Freund et al., 1997; Ozgül et al., 2011; Daiger, 2013). Often, however, individual disease genes such as *Parkin* in Parkinson's disease (Kitada et al., 1998; Brice et al., 1999) are expressed broadly in the brain (Pardo et al., 1995; Bergeron et al., 1996; Stichel et al., 2000), and expression studies would provide little guidance as to which cells are relevant to a given disease. There are also examples of clearly implicated disease genes (such as *Lrrk2* in Parkinson's disease) that are somewhere in between these two extremes—with enriched expression in striatum and a limited number of other cell types in the brain, but not ubiqui-tous expression (Galter et al., 2006; Han et al., 2008). Therefore,

A

B

Figure 1. A topography of cell-specific and enriched transcripts in the mouse brain. **A**, Hierarchical clustering of cell types by transcript levels recapitulates known biological relationships. **B**, Example of a single bullseye plot. For each cell type, the size of the bullseye in **A** is scaled to the number of specific and enriched transcripts at different stringency thresholds. For example, Purkinje neurons have many unique transcripts (large central hexagon, pSI < 0.0001)), while cortical projection neurons (in **A**) have few. For CSEAs in later figures, bullseyes will be color coded by Fisher's exact test *p* values as shown.

while a strict SE hypothesis—that disease-causing genes are always specifically expressed in the cell types disrupted by the disorder—clearly will not hold, a softer variation of the SE hypothesis might be true: across a large number of genes implicated in a disorder, we can hypothesize a relative overabundance of disease-causing genes will be found with enriched expression in the vulnerable cell types.

Modern methods have identified dozens of candidate genes that may contribute to psychiatric disorders, using common variant genome-wide association studies (GWASs) of single-nucleotide polymorphisms, rare variant analyses of single nucleotide variations (SNVs) and copy number variations (CNVs), or postmortem gene expression analysis. In previous work, we generated bacTRAP mouse lines profiling the expression of protein-coding genes in dozens of targeted cell types (Doyle et al., 2008; Dougherty et al., 2012, 2013; Dalal et al., 2013; Xu et al., 2013). Here, we combine these two sources of information and test the hypothesis

that the SE of genes can guide us toward identifying the cellular disruptions mediating disorders. We have called this approach cell type-specific expression analysis (CSEA).

Materials and Methods

Analysis of cell-specific profiling data to identify cell-specific and enriched genes. All the mouse cell type-specific gene expression profiling experiments were conducted on a single platform, most using published translating ribosome affinity purification (TRAP) data [GEO: GSE13379 (Doyle et al., 2008); GSE30626 (Dougherty et al., 2012); GSE38668 (Dalal et al., 2013); and GSE36068 (Dougherty et al., 2013], and habenula data [GSE43164 (Görlich et al., 2013)], except for retinal cell data (Corbo et al., 2007). Cell-specific gene lists for the cell types (Fig. 1) were identified by the specificity index probability (pSI) statistic at thresholds from 0.05 to 1e-4, as described previously (Dougherty et al., 2010), using Entrez CDF files version 14 (Dai et al., 2005). Data were from mice of both sexes.

Analysis of human transcriptome data identify region/time-specific and enriched genes. The human region- and time-specific gene expression

analysis experiments were conducted using published RNAseq data downloaded from Brainspan: Atlas of the Developing Human Brain (Developmental Transcriptome: RNA-Seq Gencode v3c summarized to genes; Brainspan, 2013). First, transcripts were filtered to include well annotated protein-coding genes using RefSeq (release 60) gene annotations, and to remove transcripts below a background of 0.3 reads per kilobase per million (RPKM), as described previously (Ramsköld et al., 2009).

To reduce complexity, the original 171 samples (all possible combinations of original regions and developmental periods, from either sex) were aggregated into 6 major regional divisions across 10 developmental periods, and biological replicates were averaged. Then, specific and enriched transcripts were identified by the pSI algorithm at thresholds ranging from 0.05 to 1e-4 for the following two distinct comparative analyses: first, pSI values were calculated solely for the young adulthood samples across the six brain regions. This most closely approximates the time point from the mouse data. For the second comparative analysis, pSI values were calculated for all possible comparisons of the 6 brain regions and 10 developmental periods creating a total of 60 unique combinations.

Candidate gene lists. The curated list of genes identified in human congenital retinopathies (120 genes) was downloaded from a curated database (RetNet; Daiger, 2013).

The list of candidate genes affecting human height was obtained from Lango Allen et al. (2010, their supplemental Table S13).

Lists of postmortem gene expression from human narcoleptic subjects and mice with Hcrt neuron ablation were taken from Honda et al. (2009). Mouse data were also taken from Honda et al. (2009, their Table 3), and the human genes were the nine genes replicated by quantitative PCR, as described in their text.

Autism candidate genes from the transcriptome of human brain were derived from GEO: GSE28521, analyzed as described previously (Voineagu et al., 2011). For those genes with multiple probesets, probesets were averaged, resulting in a total of 8846 genes in 58 cortex samples (29 autism, 29 controls) and 21 cerebellum samples (11 autism, 10 controls). Differential expression was assessed using the Significance Analysis of Microarrays package (Tusher et al., 2001), for a false discovery rate of <0.05 and a fold change of >1.3. A total of 463 genes were differentially expressed in cortex; 229 were downregulated (control/autism ratio, >1.3) and 234 were upregulated (autism/control ratio, >1.3; gene lists available in R package pSI). As only two genes were differentially expressed in cerebellum, we focused on the cortical data for CSEA. We also analyzed Table S5 from Chow et al., 2012 and Table S1 from Garbett et al., 2008.

A curated list of autism candidate genes derived from human genetics studies was downloaded from the AutDB (Autism Database; Basu et al., 2009) before the addition of data from the exome studies. The data were for 328 genes, including rare single gene variants, disruptions/mutations, deletions/duplications directly linked to autism; genes implicated in syndromic autism; small risk-conferring common variants identified from association studies; and functional candidate genes not yet experimentally linked with autism.

Lists of *de novo* variants were curated from recent autism exome sequencing studies (Iossifov et al., 2012; Neale et al., 2012; O'Roak et al., 2012; Sanders et al., 2012), including 122 protein-disrupting and 528 mis-sense mutations.

SEA statistical testing. To evaluate the significance of overlap between a candidate gene list and the list of transcripts enriched in a particular cell type, we used the Fisher's exact test (Fisher, 1922), with Benjamini–Hochberg multiple testing correction for the number of cell types assayed (Benjamini and Hochberg, 1995). Though gene lists at different pSI thresholds for the same cell type are clearly not independent tests, using a more conservative threshold that assumes they are, performing a Benjamini–Hochberg corrected for all cell types and thresholds did not substantially alter any of the results of this article.

For human RNAseq data compared with human gene lists, the background was set as any protein-coding gene with RPKM > 0.3, a threshold indicating detectable expression (Ramsköld et al., 2009). For cross-species comparisons, the "background" gene list for these analyses is made by the intersection of the following two lists: (1) all genes on the mouse microarray; and (2) all genes with clear mouse–human ho-

mologs. We mapped human-to-mouse homologs based on the UCSC (University of California Santa Cruz) Genome Browser table "kgxref," accessed on May 15, 2013.

Sensitivity and specificity analyses. The retinopathy disease genes ($n = 120$) were used to investigate the impact of candidate gene list size and purity on the sensitivity and specificity of CSEA.

To model the impact of list size, we tested gene lists of 13 sizes ($n = 1$, 10, 20, …120). For each, we generated a sampled candidate gene list 100 times by randomly selecting genes from the retinopathy gene list. These 100 lists were then analyzed by CSEA across all cell types and pSI thresholds. To get a sense of the false-positive frequency, we also counted the number of times that a nonretinal cell type was falsely implicated ($p < 0.05$) across all samples.

To model the impact of candidate gene list purity (i.e., the impact of including spurious candidate genes), we tested gene lists of fixed size ($n = 120$) with 11 different proportions of retinopathy genes. Specifically, in a sampled list, 0%, 10%, …100% of its genes were randomly selected from the retinopathy gene list, while the remaining genes were randomly selected from the genome. For each purity level, we sampled and analyzed the genes as indicated above.

Results
CSEA

First, we merged the data from our surveys of cell type-specific gene expression from the mouse brain (Doyle et al., 2008; Dougherty et al., 2012, 2013; Dalal et al., 2013; Görlich et al., 2013). For the purposes of the specific hypotheses cited below, we have also incorporated published data profiling specific retinal cell types on the same platform (Corbo et al., 2007). Next, we analyzed these profiles with our validated analytical pipeline for detecting cell type-specific and enriched genes (Dougherty et al., 2010). This approach compares each cell profile to all other profiles and identifies transcripts consistently enriched in each cell type, calculates an enrichment score (the SI), and ascribes a pSI value. Cell-specific and enriched transcript lists can then be derived for each cell type at a given pSI threshold. The lower the pSI, the smaller, yet more stringently specific, the transcript list will be (Dougherty et al., 2010). As the softer SE hypothesis does not predicate any particular enrichment threshold, we varied pSI systematically throughout our analyses. Hierarchical clustering of cell types using genes with significant pSI values in any cell type recapitulates the ontogeny of these cells—there are large branches that separate glia from neurons, and related subtypes of neurons are loosely clustered together (Fig. 1A). As expected, markedly distinct cell types, such as rods, have larger numbers of uniquely expressed transcripts at any pSI threshold (gene list size is proportional to the area of the polygons; Fig. 1B), while less distinct cell types such as cortical projection neurons have smaller numbers of specific transcripts, especially at stringent pSI thresholds. Though inherently relative, this clustering, and the composition of the gene lists, are relatively stable. For example, even dropping one of two closely related astrocyte samples (astrocytes of the cerebellum) and repeating the pSI analysis has no impact on the overall clustering of the remaining samples, and results only in subtle changes in the gene lists themselves (Fig. 2A,B). Furthermore, examining the transcript length distributions we confirmed that there was no relationship to pSI at any threshold (Fig. 2C), confirming that transcript length would not bias our results. Having identified sets of transcripts enriched in each cell type, we then mined available human data from brain transcriptomic studies and catalogs of disease-affiliated genetic variants to generate lists of candidate genes. There is a wide variety of potential sources for candidate gene lists. A priori, we can assume that certain features of the candidate gene list will influence the outcome. First, given that the softer variation of the SE hypothesis proposes only a mild statistical enrichment of risk genes in the cell type of interest, the method will likely require a fairly large number of
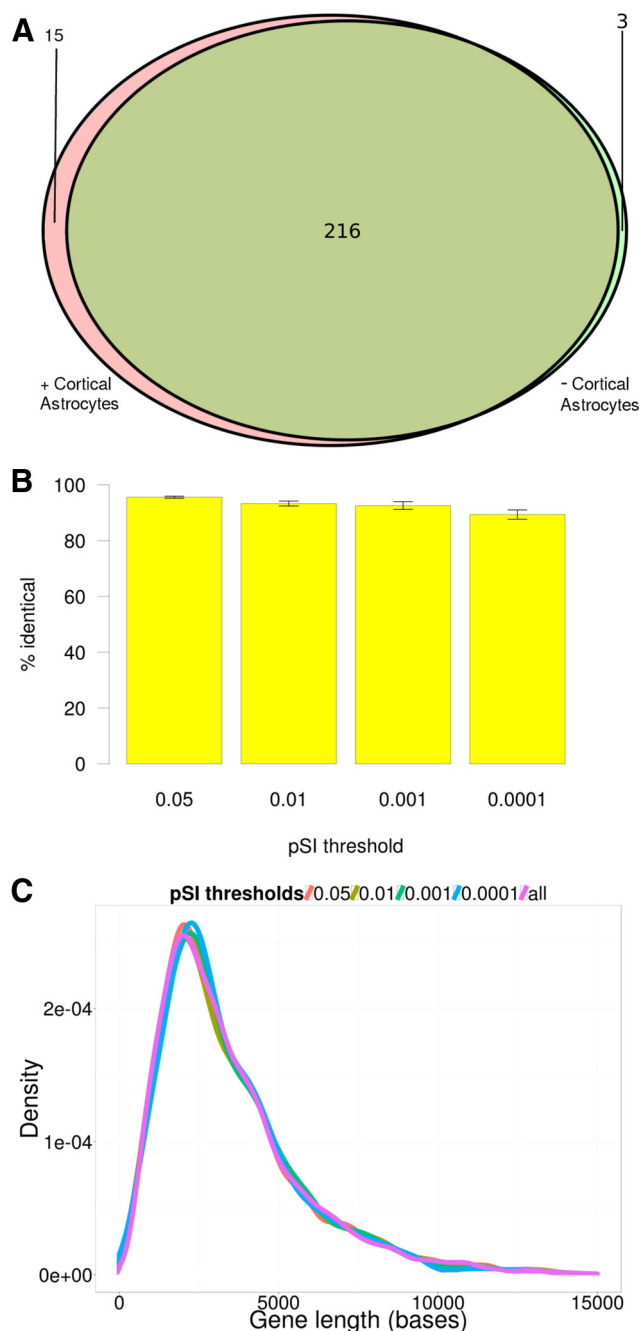
Figure 2. The cell-specific and enriched transcript lists are robust and unrelated to transcript length. **A**, Venn diagram showing little change in transcript lists for cortical astrocytes at pSI = 0.01, when pSI is recalculated with or without the closely related cerebellar astrocyte sample. **B**, Across all cell types and pSI thresholds, gene lists are >89% identical when calculated with or without cerebellar astrocytes. **C**, Density plot of distributions of transcript lengths colored by pSI value shows no length bias.

genes to detect a significant signal, especially if multiple cell types are disrupted in a disorder. Second, many methods of identifying candidate regions of the genome do not perfectly indicate individual genes—both GWASs and CNV studies may implicate regions containing many genes, any of which could be the causative gene or genes. This uncertainty at the level of candidate genes could also detract from the ability of our method to identify cell types. Therefore, the ideal candidate gene list would contain a reasonably large number of genes, each of which was robustly associated with the phenotype of interest. As positive controls, we identified one such

candidate gene list from human genetics (genes causing familial retinopathies), and one from postmortem gene expression data (narcolepsy). We also tested a negative control candidate gene list (loci associated with height) to get a sense of the false-positive rate of our method. We also investigated the impacts of candidate gene list size and purity on the sensitivity and specificity of the CSEA before examining our test case: autism candidate genes from various sources.

CSEA of retinopathies
There are now >100 genes that are clearly associated with congenital blindness and other disorders of the retina (Daiger, 2013). Most of these genes have classic Mendelian inheritance patterns with high penetrance, and thus serve as an ideal positive control in terms of gene number and strength of association. If our hypothesis is correct, then these genes should be over-represented in expression profiles of the primary cell types of the eye—rods or cones. Applying the CSEA, we detected a highly significant enrichment of retinopathy genes in the profiles of rod- and cone-enriched genes (Fig. 3A). This result provides support for the SE hypothesis and proof-of-principle for our approach. If we had not known retinopathies were diseases of rods and cones, this analysis would have very clearly indicated that they were the relevant cell types.

CSEA of height
The GIANT consortium has now identified hundreds of loci that contribute to the phenotype of height across individuals (Lango Allen et al., 2010). For the purpose of the CSEA, we make the reasonable assumption that it is unlikely there is a particular cell type in the brain responsible for height. Therefore, any signal seen in this analysis should give us a sense of the magnitude and number of false positives that a list of candidate genes might provide. Analysis with these candidate genes reveals no signals (figure not shown).

CSEA of transcriptome data in narcolepsy
In the analyses above, we tested the CSEA method on gene lists derived from human genetics studies. However, the method could equally apply to postmortem transcriptomic data to permit cellular inferences regarding the consequences of disease. For example, if a disease results in the loss of a particular cell type, the transcripts specific to that cell type should disappear. To test this approach, we combined our analysis with a transcriptomic study of narcolepsy (Honda et al., 2009). Narcolepsy can be caused by a loss of Hcrt-expressing neurons (Peyron et al., 2000), which we have profiled (Dalal et al., 2013). Microarray data from mice where these neurons have been ablated identified 63 downregulated genes, which were sufficient to specifically indicate a loss of Hcrt neurons (Fig. 3B). In human patients, even though only nine transcripts were replicated as decreased, these likewise indicated that narcolepsy is a pathology of Hcrt neurons (Fig. 3B).

Sensitivity and specificity analysis of CSEA
Unlike retinopathies, most complex psychiatric disorders have only a small number of strongly associated loci. In addition, several of the current methodologies for identifying loci do not provide clear evidence as to which genes in the loci actually contribute to risk. Thus, in psychiatric genetics, we might have both fewer candidates and potentially lower confidence in each of the candidates. Therefore, we used the retinopathy data to determine the impact of varying these parameters on the sensitivity and specificity of CSEA.

To model lower candidate gene numbers, we sampled random subsets of the retinopathy gene list to generate candidate lists of
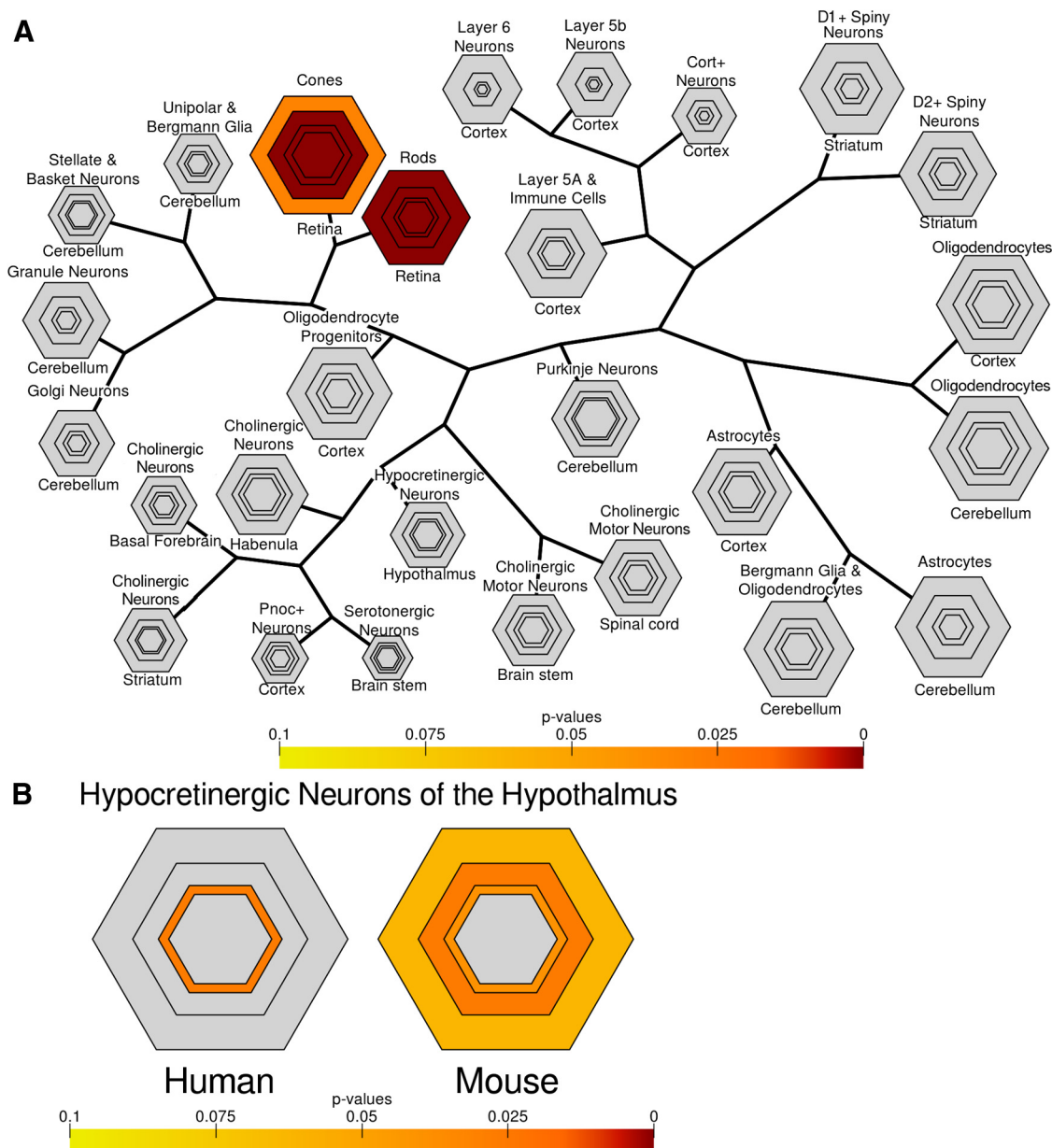
Figure 3. CSEA correctly identifies retinopathies as diseases of rods and cones, and CSEA of narcoleptic transcriptomic data identifies a loss of hypocretin neurons. ***A***, Bullseye plot of the output of CSEA reveals a substantial over-representation of retinopathy disease genes ($n = 120$) producing transcripts enriched in rods and cones, regardless of the threshold chosen for pSI. ***B***, Output of CSEA reveals an over-representation of hypocretin neuron cell transcripts among those transcripts ($n = 9$) that were decreased in the hypothalami of human narcoleptic subjects or in mice with Hcrt neuron ablation ($n = 63$).

varying numbers. Figure 4A shows the impact of list size on the average p value for the overlap with rods. It is evident that the number of candidate genes has a substantial impact on the sensitivity of the method. With very small numbers of genes, we are unlikely to detect enrichment in rods after multiple testing corrections. However, when the number of genes was ≥30, >94% of the sampled gene lists would have correctly implicated rods in retinopathies at any pSI threshold. A pSI threshold of <0.01 is shown here. Results are substantially similar across all pSI thresholds from 0.05 to 0.0001 (data not shown).

To model the impact of including spurious candidate genes, we replace random subsets of the retinopathy gene list with an equivalent number of random genes (Fig. 4B). As the purity of the gene list decreased, the enrichment signal in the rods gradually decreased. Yet, overall, the method was remarkably robust; even

when only 30% of the genes on the list were true retinopathy genes, >90% of the time the CSEA still implicated rods. Again, a threshold of pSI < 0.01 is shown here, and results are substantially similar across all pSI thresholds (data not shown).

In the same analysis, we also counted how frequently a non-retinal cell type was falsely implicated. Varying purity as above, for a pSI threshold of 0.05, of 1100 total samplings at any purity, a nonretinal cell type was implicated 56 times, suggesting a false-positive rate of ∼5%. There was no clear relationship between purity and false-positive rate. With any pSI < 0.01, this false-positive rate dropped to <1%. Varying the size of the gene list, as above, had no impact on the false-positive rate, with <1% of samples showing false-positive results. Overall, while larger, lower stringency lists (pSI < 0.05) are slightly more vulnerable to false-positive results, our analysis suggests that, with appropriate
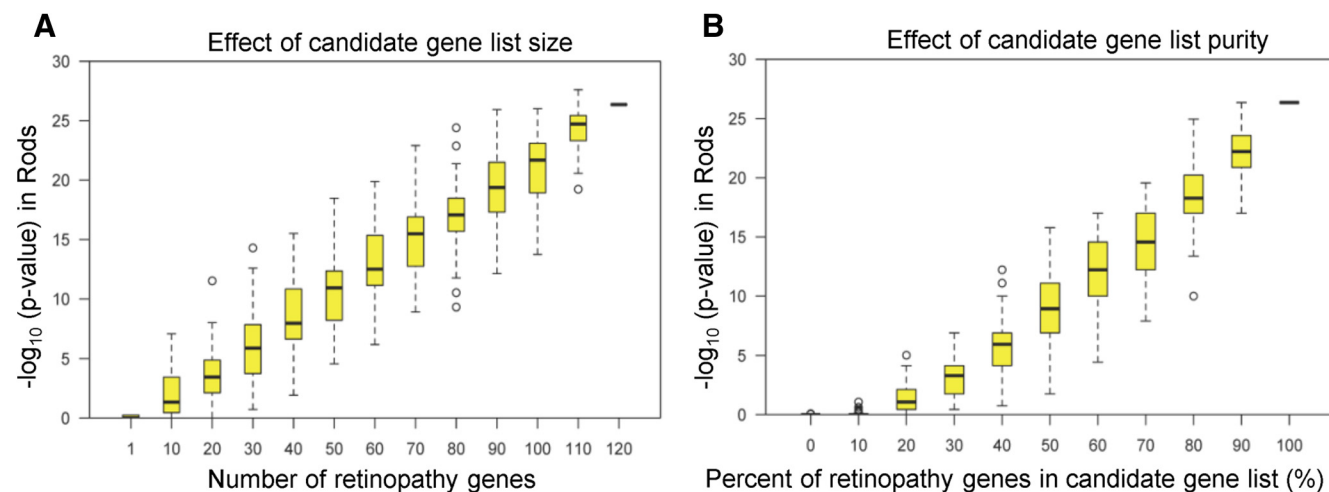
Figure 4. Impact of candidate gene list size and purity on sensitivity and specificity of CSEA. ***A***, Box-and-whisker plots illustrating the distribution of CSEA *p* values (log base 10 scale, *y*-axis) for rods from retinopathy gene lists sampled at various sizes (100 samplings per size). ***B***, From random subsets of retinopathy gene lists mixed with random sets of nonretinopathy genes. The *x*-axis (purity) indicates what fraction of the list derives from the original retinopathy set (100 samplings per purity). A pSI threshold of <0.01 is shown here. Results are substantially similar across all pSI thresholds from 0.05 to 0.0001 (data not shown).

multiple testing corrections, the overall false-positive rate is acceptably low. This also suggests that significant results for the same cell type across multiple thresholds should be given more credence than those that are only identified at a single pSI threshold.

In all, our analysis suggests that, with the emerging candidate gene numbers for psychiatric disorders, we should have sufficient power to detect a signal in some cell types.

Test case: autism
Unlike retinopathies or narcolepsy, the disrupted cell types for autism are not known. Is autism a disease of a single cell type? Or are there multiple cellular mechanisms contributing to the development of the disorder? The field is now accumulating reasonable sets of candidate genes for this disorder, both from studying gene expression in postmortem brain and from human genetics studies. Based on the analysis above, the size of these gene lists finally justifies a comprehensive analysis. Therefore, we sought to determine the following: (1) whether SE may identify cell types relevant to mechanisms of this disorder; (2) whether these consist of a single cell type or multiple cell types; and (3) whether these results were reproducible across multiple studies and data types.

Transcriptome profiling
A recent tour-de-force gene expression study screened postmortem cortical tissue from 29 individuals with autism and matched controls (Voineagu et al., 2011). Alterations of gene expression from a complex mix of cells such as those from cortex may represent changes in the relative cellular composition of the tissue (Shen-Orr et al., 2010; Kuhn et al., 2011; Xu et al., 2013). The expression results from this human postmortem study were applied to CSEA on specific genes in cortex cell types to determine whether they predict changes in particular cell types. Here there were clear signals (Fig. 5). Upregulated genes were overrepresented in astrocytes and immune cell types, while downregulated genes were over-represented in the profiles from presumptive neurons (especially Pnoc+ neurons, which are largely interneurons). These results held across multiple pSI thresholds, and after multiple testing corrections. The most parsimonious interpretation of these results is that there is a relative loss of Pnoc+ cells from these autistic brain samples, and a cor-

responding increase in astrocyte and immune cell (potentially microglial) numbers.

Two other studies provide some support for glial deficits in autism transcriptome profiles. Examining the results of an earlier study in cortex with a smaller number of subjects (Garbett et al., 2008) replicates the enrichment of astrocyte genes among those upregulated (*n* = 130) in autism, though few genes were detected as downregulated genes (*n* = 22), and these did not map to a particular cell type (figures not shown). There may be some age dependence to these effects, as the data from Chow et al. (2012) show a downregulation of oligodendrocyte genes from youths in autism (figure not shown), but no overlap with astrocytes. The discrepancies seen between the three transcriptomic datasets are independent of the CSEA method, as there is little concordance in the primary data between these studies. This suggests that autism may have substantial heterogeneity not only in its genetics, but also in its cellular mechanisms, and that the different studies have captured subsets of patients with different cellular disruptions.
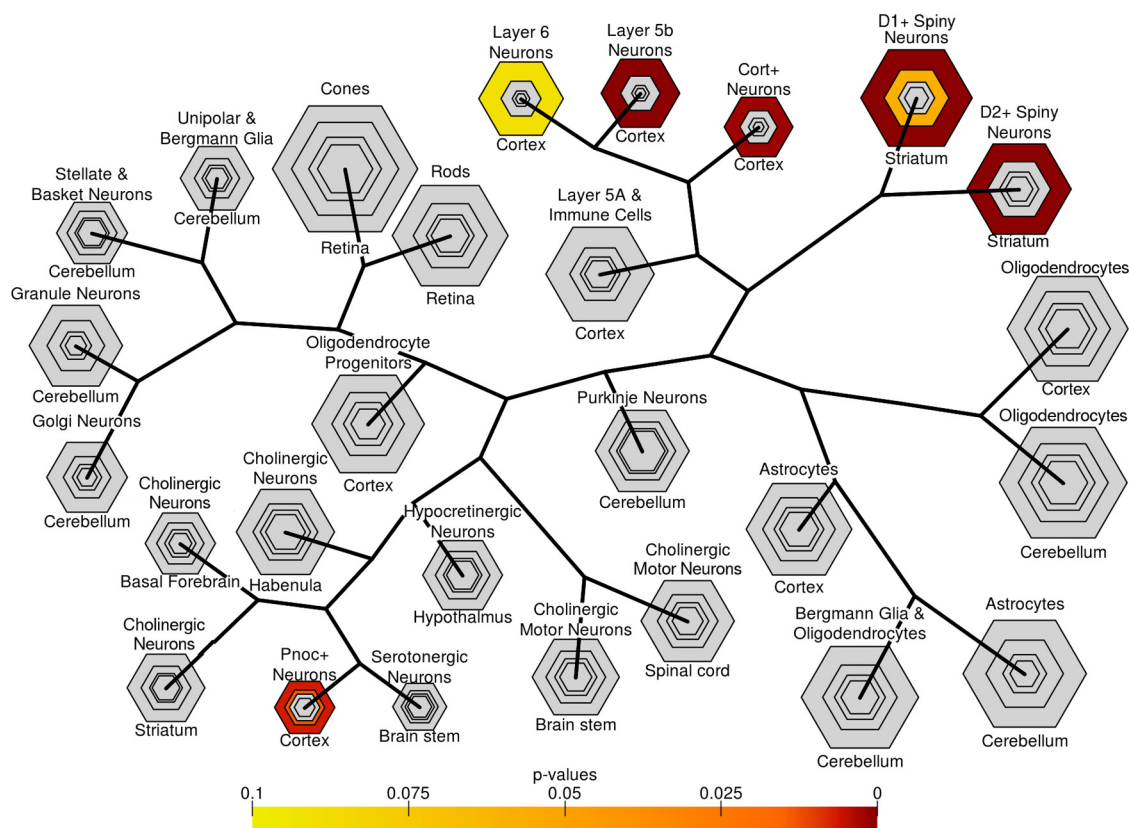
Human genetics
Many parallel approaches (Bill and Geschwind, 2009; Glessner et al., 2009; Wang et al., 2009; Pinto et al., 2010; O'Roak et al., 2012) have implicated hundreds of candidate genes and regions for autism, though very few individual genes with the strength of association seen in familial retinopathies (Abrahams and Geschwind, 2008; Devlin and Scherer, 2012). Rather than focus on the results of a single study, we mined a curated database (AutDB) for a list of candidate loci from across studies (Basu et al., 2009). Though the signals here are weaker (Fig. 6), candidate genes from human genetics are over-represented again in cortical interneurons (Pnoc+, Cort+), as well as in corticothalamic neurons (Ntsr+) and striatal medium spiny neurons.

Finally, recent exome studies have identified seven genes (SCN1A, LAMC3, GRIN2B, SCN2A1, NTNG1, KATNAL2, and CHD8) in which *de novo* variants are likely causative of autism (Iossifov et al., 2012; O'Roak et al., 2012; Sanders et al., 2012), though each gene so far has been detected in only a very small number of individuals. These mutations collectively are not enriched in any of our cell types, nor are the larger number of *de novo* proteins disrupting (*n* = 122, frame-shift, stop codon, or splice site mutations) collected across all of these studies—
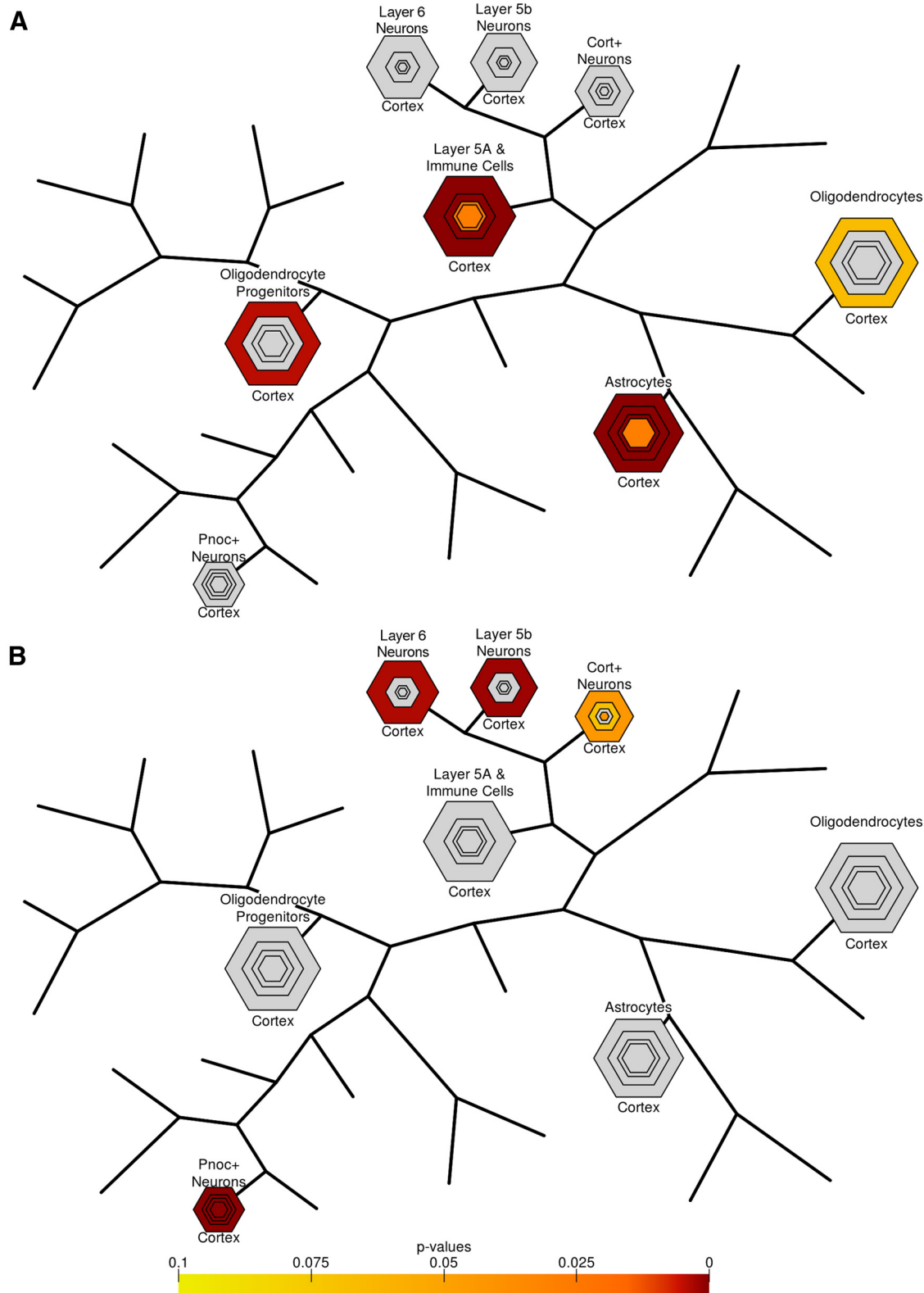
Figure 5. CSEA of autism cortical transcriptomic data suggest increased signals from astrocytes and immune cells, and a disruption of interneurons. ***A***, Output of CSEA reveals a substantial over-representation at multiple pSI levels for astrocyte and immune cell transcripts among those transcripts found to be increased ($n = 184$) in cortices from individuals with autism. ***B***, Output of CSEA reveals a substantial over-representation of Pnoc + interneuron cell transcripts at multiple pSI levels among those transcripts ($n = 174$) found to be decreased in cortices from individuals with autism (axes are as shown in Fig. 1; analysis restricted to cortical cell types).

Figure 6. CSEA of autism candidate genes curated from human genetic studies suggests a disruption of cortical neurons and striatal circuitry. Output of CSEA reveals a substantial over-representation at multiple pSI levels for cortical interneurons (Pnoc+ and Cort+), layer 5b neurons, and striatal medium spiny neurons.

though each of these individually are not yet clearly statistically associated with autism. The lack of signal compared with the AutDB suggests that perhaps there is something different about the cellular mechanisms of these *de novo* variants compared with those found in earlier autism genetics studies. In particular, many of the newly discovered autism-associated variants map to genes previously implicated in early brain development, but the available CSEA data come from adult brains. Therefore, we also tested our method using pSI values derived from human brain regions across development, calculated using the Brainspan (2013) resource.

Region- and time-specific expression analysis

While, there is no collection of human expression data with cellular resolution comparable to what has been done in the mouse, repeating the analysis with human data across development permits the identification of genes transiently expressed at certain developmental epochs, and repeating the analysis across adult regions allows an independent validation of the mouse results, at least at a crude regional level. Therefore, we condensed the Brainspan (2013) data into 6 major regional divisions across 10 developmental times, and calculated pSI values for both a comparative analysis of just the young adult regions (analogous to our mouse data; Fig. 7A), and a matrix of regions and times covering all development times (Fig. 7B).

First, examining the enrichment of autism candidate genes from the AutDB across human brain regional gene expression confirms that these genes are enriched in the cortex and striatum (Fig. 7A). This independently substantiates our mouse findings. Second, the examination of the *de novo* events from probands (Fig. 7B) across both space and time implicates a window in early

to mid-fetal development when these genes are enriched in the developing cortex, a finding similar to other recent analyses using distinct methods (Parikshak et al., 2013; Willsey et al., 2013). This again confirms the potential importance of this structure for the manifestation of autism and suggests that for some patients the circuit deficits leading to symptomatology are a consequence of malformations during fetal development. A control analysis with the 122 silent *de novo* variants from unaffected siblings from these same studies shows no mapping to any particular cell type, region, or time of development.

Discussion

Comparison to other approaches

There is a clear challenge to understanding the commonalities of sets of genes arising from high-throughput screens such as RNAseq or GWAS. A variety of consortia, notably Gene Ontologies (GO), have attempted systematic functional annotations of all genes, and a variety of software are available for detecting biological processes from these annotations that may be enriched in particular sets of candidate genes (Dennis et al., 2003; Maere et al., 2005; Shi and Walker, 2007). The hope is that these biological processes will indicate a commonality of mechanism and suggest standardized routes to treatments. In autism, these analyses have variously identified a preponderance of cell–cell adhesion molecules, GTPase signaling, ubiquitination, synaptic regulators, and neural development among candidate genes (Bill and Geschwind, 2009; Glessner et al., 2009; Wang et al., 2009; Pinto et al., 2010; O'Roak et al., 2012). However, these GO-based approaches have two key disadvantages. First, novel genes will be unannotated, and known genes will be annotated only for their known functions. Given the complexity of biology, current annotations

are likely to reflect only a modest subset of the true roles of most genes. Second, biological processes frequently do not clearly suggest potential routes to treatments. For example, if autism is a disorder of cell–cell interaction, which cells, and which interactions, would need to be targeted for treatment?

Here, we have outlined and tested an alternative approach (CSEA) focused on identifying candidate cell types that may be disrupted by human neurological disorders. In many ways, our approach is analogous to GO analyses, though we used information derived from gene expression rather than gene annotation. Indeed, we view our method as an important new complement to those existing tools. Compared with these previous methods, our method has the advantages that nothing needs to be known about the function of the genes a priori, and that knowledge of disrupted cell types may suggest more obvious routes to treatment. It has the disadvantage that it will be able to detect the signal only for those cell types, and developmental stages, that have been profiled. While clearly the method will become more informative as additional cell types are incorporated, particularly with the development of new TRAP reporter mice (Hupe et al., 2013; Zhou et al., 2013), here we have partially compensated for the incompleteness of our current survey by including profiles from brain regions as placeholders for the cells within them. While this may be an imperfect proxy, this disadvantage will gradually resolve as more cell types are profiled. It is also important to note that any measure of specificity is relative, and the addition of new tissues or cell types will slightly alter the exact composition of a cell-specific list. Nonetheless, the collection of cells currently included appears relatively robust to minor perturbations of this manner. Repeating the analyses for all figures on a pSI dataset calculated with one cell type excluded (Fig. 2A,B) did not alter which cell types were detected by CSEA (figure not shown).

For human expression data, there are other methods, notably weighted gene co-expression network analysis (WGCNA), that have been used to identify what might be the signatures of particular cell types (Langfelder and Horvath, 2008; Oldham et al., 2008). These methods have the advantage that they do not depend on external information (they are more "bottom-up"); the signal is derived from correlations within the dataset itself and can allow inferences from human sources (Kang et al., 2011) across a variety of ages. However, the measures of cell type expression are indirect, of limited sensitivity for very rare cell types, and will not be able to dissociate tightly correlated cell types. Also, these analyses can only be applied to expression data, and not human genetics studies. Nonetheless, based on this bottom-up approach, the authors of one of the transcriptomic studies (Voineagu et al., 2011) interpreted their results largely in the same
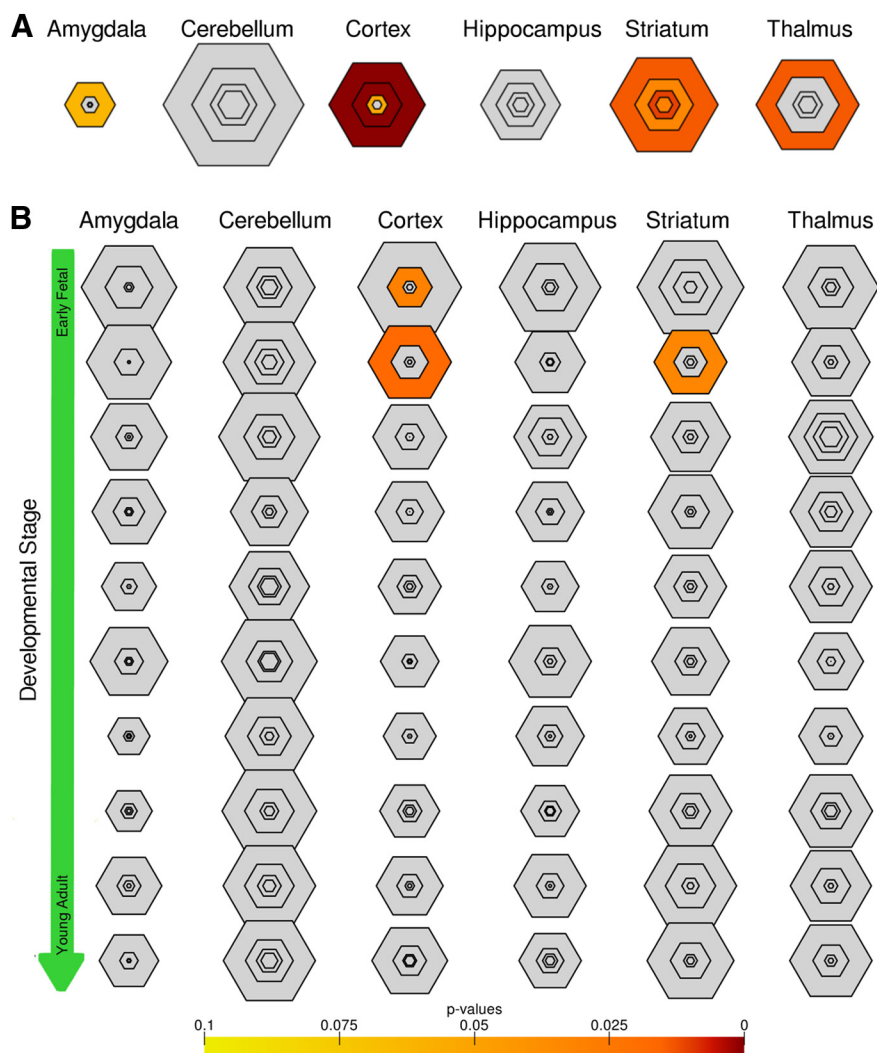
Figure 7. Analysis of enrichment in human brain regions supports the disruption of cortex and striatum. **A,** Curated autism genes are over-represented at multiple pSI levels in striatally and cortically enriched transcripts calculated from Brainspan, 2013 young adult RNAseq data. **B,** Rare, protein-disrupting *de novo* variants identified in probands with autism (*n* = 122) have enriched expression during fetal cortical and striatal development.

way that we would from our more directed approach: a relative decrease in interneuron signal, and a corresponding increase in glial and immune-related genes. Joint analyses using our approach with WGCNA can further improve the power to detect alterations of cellularity from transcriptomic data (Xu et al., 2013).

Finally, as cell-specific profiling is not as mature in human samples, our method sometimes depends on an assumption of a degree of conservation between human and mouse gene expression. Systematic evaluation of laminar-specific genes in cortices of both species suggests good conservation of expression for ~80% of genes (Zeng et al., 2012). Nonetheless, we have also included pSI values for human brain region data to permit at least region-level analyses directly from human expression data.

Is autism a disease of a particular cell type?
Retinopathies are clearly a disease of rods and cones, and the strong signals seen in Figure 3 demonstrate that, at least for these cell types, our soft SE hypothesis holds. We also have identified the loss of Hcrt neurons from narcoleptic brains, even with fairly small numbers of gene expression changes. Thus, CSEA can identify cellular mechanisms for neurogenetic disorders under some circumstances.

Our results for autism are somewhat less consistent across data sources. Two of three transcriptional studies suggest there is an increase in glia in the brains of individuals with autism, perhaps reflecting some kind of gliosis, with potentially an autoimmune component. This would be consistent with recent work suggesting a significant preponderance of anti-brain antibodies in the serum of mothers of children with autism (Brimberg et al., 2013), as well as microglial activation in the brains of young adults with autism (Suzuki et al., 2013).

It is also very interesting to note that both a transcriptomic study and the human genetics database provide some evidence implicating the cortical interneuron. This supports proposals that one of the cellular mechanisms of autism may be a relative deficit in the function or presence of interneurons (Marín, 2012). This is consistent with mouse research demonstrating that interneuron-specific knockouts of even broadly expressed syndromic autism genes (such as the *Mecp2* gene responsible for Rett syndrome) can result in autism-like phenotypes (Chao et al., 2010; Saunders et al., 2013), and complete knockout of other autism candidate genes can result in a specific loss of interneurons (Sgadò et al., 2013).
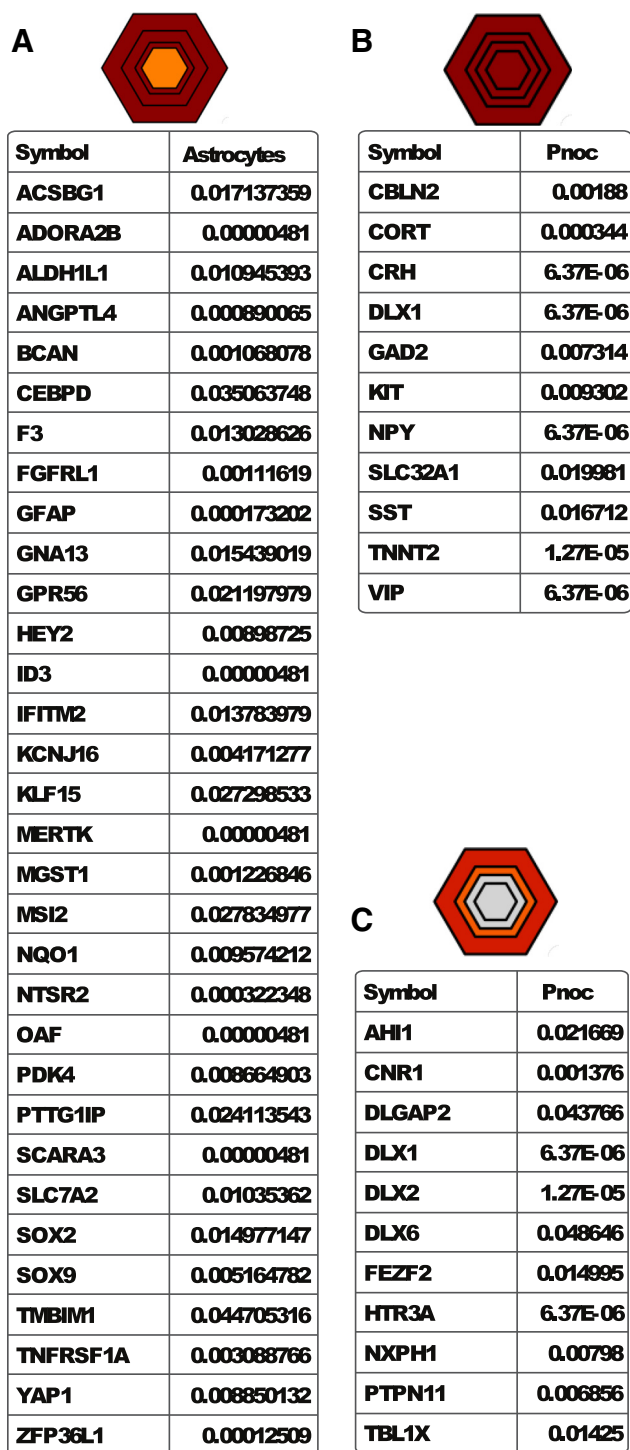
It is expected there will be some divergence between the human genetics findings and the gene expression studies. Compared with human genetics, gene expression is likely to show signals derived from both the causes and the consequences of the disease. For example, while a signal from genes found in interneurons was decreased in autism, there was a relative increase in the signals from astrocytes and putative immune cells. This could represent the attempt of the nervous system to support or remove dysfunctional cells, or might posit some kind of inflammation-mediated mechanism for the disruption of interneurons in some patients (Wills et al., 2011). Consistent with this interpretation, it is interesting to note that the genes driving the implication of Pnoc+ neurons are different between the genetic (Fig. 8C) and transcriptomic approaches (Fig. 8B). Likewise, the astrocyte-implicated genes include at least two validated markers of these cells, *Gfap* and *Aldh1l1* (Fig. 8A), which have no link to the genetics of autism.

There are several cell types and regions long considered as autism candidates for which we see no signal in any of our autism analyses, for example, the cerebellum or serotonergic neurons (Dougherty et al., 2013). While CSEA cannot be used to definitively exclude particular cell types, this suggests that disruptions acting specifically in these cells might be having their final consequences on downstream cortical or striatal cell types, both of which are structures innervated by serotonergic cells.

In total, the application of CSEA to autism suggests that autism is not a disorder of just a single cell type. Across different data sources, different, but somewhat overlapping, cell types were implicated, with signals in astrocytes, immune cells, cortical neurons (especially interneurons), and striatal medium spiny cells. These observations suggest that there are potentially multiple cellular routes to autism, though it is unclear whether multiple cell types need to be disrupted in each individual or whether these routes occur in different subsets of patients. This is an interesting area for future investigation.

R package and on-line tool

To further facilitate the distribution of our data and approach into the wide range of GO and GO-like tools available, we have packaged our pSI code as an R package freely distributed from the CRAN repository, with several useful functions for calculating pSI, calculating Fisher's exact tests, and returning the intersect

A

Symbol	Astrocytes
ACSBG1	0.017137359
ADORA2B	0.00000481
ALDH1L1	0.010945393
ANGPTL4	0.000890065
BCAN	0.001068078
CEBPD	0.035063748
F3	0.013028626
FGFRL1	0.00111619
GFAP	0.000173202
GNA13	0.015439019
GPR56	0.021197979
HEY2	0.00898725
ID3	0.00000481
IFITM2	0.013783979
KCNJ16	0.004171277
KLF15	0.027298533
MERTK	0.00000481
MGST1	0.001226846
MSI2	0.027834977
NQO1	0.009574212
NTSR2	0.000322348
OAF	0.00000481
PDK4	0.008664903
PTTG1IP	0.024113543
SCARA3	0.00000481
SLC7A2	0.01035362
SOX2	0.014977147
SOX9	0.005164782
TMBIM1	0.044705316
TNFRSF1A	0.003088766
YAP1	0.008850132
ZFP36L1	0.00012509

B

Symbol	Pnoc
CBLN2	0.00188
CORT	0.000344
CRH	6.37E-06
DLX1	6.37E-06
GAD2	0.007314
KIT	0.009302
NPY	6.37E-06
SLC32A1	0.019981
SST	0.016712
TNNT2	1.27E-05
VIP	6.37E-06

C

Symbol	Pnoc
AHI1	0.021669
CNR1	0.001376
DLGAP2	0.043766
DLX1	6.37E-06
DLX2	1.27E-05
DLX6	0.048646
FEZF2	0.014995
HTR3A	6.37E-06
NXPH1	0.00798
PTPN11	0.006856
TBL1X	0.01425

Figure 8. Autism genes and transcripts enriched in particular cell types. ***A***, Transcriptionally upregulated transcripts that overlap with astrocyte-enriched genes, corresponding to Figure 5A. Right column, pSI in cortical astrocytes. ***B***, Transcriptionally downregulated transcripts that overlap with Pnoc + neuron-enriched transcripts, corresponding to Figure 5B. ***C***, Autism candidate genes that overlap with Pnoc + neuron-enriched transcripts, corresponding to Figure 6.

between certain cell types and candidate gene lists. Most importantly, we have included the data and candidate gene lists used in this manuscript, including our cell type data, for the end user. This package should speed the propagation of our general data and this style of approach to other tool developers. It will permit the more savvy users to readily implement any variation in our analyses that they would like.

While we tested our CSEA approach with autism candidate genes, this approach could be applied to other disorders of the nervous system, as well as to other cutting edge sources of transcriptomic data where there is a need for methods to identify cellular composition from the data (Knight et al., 2012). Therefore, in addition to the R package, we have also provided a web server (http://genetics.wustl.edu/jdlab/csea-tool-2/) that will permit other investigators to easily examine potential SE patterns across mouse cell types, human adult brain regions, or human brain development using their own candidate gene lists. We hope that the identification of candidate cell types from the genetic information will suggest novel routes to treatments in these disorders.

References

Abrahams BS, Geschwind DH (2008) Advances in autism genetics: on the threshold of a new neurobiology. Nat Rev Genet 9:341–355. CrossRef Medline

Basu SN, Kollu R, Banerjee-Basu S (2009) AutDB: a gene reference resource for autism research. Nucleic Acids Res 37:D832–D836. CrossRef Medline

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—a practical and powerful approach to multiple testing. J R Stat Soc Ser B 57:289–300.

Bergeron C, Petrunka C, Weyer L (1996) Copper/zinc superoxide dismutase expression in the human central nervous system— correlation with selective neuronal vulnerability. Am J Pathol 148:273–279. Medline

Bill BR, Geschwind DH (2009) Genetic advances in autism: heterogeneity and convergence on shared pathways. Curr Opin Genet Dev 19:271–278. CrossRef Medline

Brainspan (2013) BrainSpan atlas of the developing human brain Available at www://brainspan.org.

Brice A, Lucking CB, Abbas N, Durr A, Ricard S, Bonifati V, De Michele G, Wood NW, Gasser T, Breteler M, Harhangi S, Oostra B, Filla A, Meco G, Denefle P, Agid Y (1999) Mutations in the parkin gene are a common cause for autosomal recessive early onset parkinsonism in Europe. Neurology 52:A554–A555.

Brimberg L, Sadiq A, Gregersen PK, Diamond B (2013) Brain-reactive IgG correlates with autoimmunity in mothers of a child with an autism spectrum disorder. Mol Psychiatry 18:1171–1177. CrossRef Medline

Cahoy JD, Emery B, Kaushal A, Foo LC, Zamanian JL, Christopherson KS, Xing Y, Lubischer JL, Krieg PA, Krupenko SA, Thompson WJ, Barres BA (2008) A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J Neurosci 28:264–278. CrossRef Medline

Chao HT, Chen H, Samaco RC, Xue M, Chahrour M, Yoo J, Neul JL, Gong S, Lu HC, Heintz N, Ekker M, Rubenstein JL, Noebels JL, Rosenmund C, Zoghbi HY (2010) Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. Nature 468:263–269. CrossRef Medline

Chow ML, Pramparo T, Winn ME, Barnes CC, Li HR, Weiss L, Fan JB, Murray S, April C, Belinson H, Fu XD, Wynshaw-Boris A, Schork NJ, Courchesne E (2012) Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature ages. PLoS Genet 8:e1002592. CrossRef Medline

Corbo JC, Myers CA, Lawrence KA, Jadhav AP, Cepko CL (2007) A typology of photoreceptor gene expression patterns in the mouse. Proc Natl Acad Sci U S A 104:12069–12074. CrossRef Medline

Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H, Watson SJ, Meng F (2005) Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res 33:e175. CrossRef Medline

Daiger SP (2013) RetNet, the Retinal Information Network. https://sph.uth.edu/retnet/.

Dalal J, Roh JH, Maloney SE, Akuffo A, Shah S, Yuan H, Wamsley B, Jones WB, Strong Cde G, Gray PA, Holtzman DM, Heintz N, Dougherty JD (2013) Translational profiling of hypocretin neurons identifies candidate molecules for sleep regulation. Genes Dev 27:565–578. CrossRef Medline

Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 4:P3. CrossRef Medline

Devlin B, Scherer SW (2012) Genetic architecture in autism spectrum disorder. Curr Opin Genet Dev 22:229–237. CrossRef Medline

Dougherty JD, Schmidt EF, Nakajima M, Heintz N (2010) Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. Nucleic Acids Res 38:4218–4230. CrossRef Medline

Dougherty JD, Fomchenko EI, Akuffo AA, Schmidt E, Helmy KY, Bazzoli E, Brennan CW, Holland EC, Milosevic A (2012) Candidate pathways for promoting differentiation or quiescence of oligodendrocyte progenitor-like cells in glioma. Cancer Res 72:4856–4868. CrossRef Medline

Dougherty JD, Maloney SE, Wozniak DF, Rieger MA, Sonnenblick L, Coppola G, Mahieu NG, Zhang J, Cai J, Patti GJ, Abrahams BS, Geschwind DH, Heintz N (2013) The disruption of Celf6, a gene identified by translational profiling of serotonergic neurons, results in autism-related behaviors. J Neurosci 33:2732–2753. CrossRef Medline

Doyle JP, Dougherty JD, Heiman M, Schmidt EF, Stevens TR, Ma G, Bupp S, Shrestha P, Shah RD, Doughty ML, Gong S, Greengard P, Heintz N (2008) Application of a translational profiling approach for the comparative analysis of CNS cell types. Cell 135:749–762. CrossRef Medline

Ercan-Sencicek AG, Stillman AA, Ghosh AK, Bilguvar K, O'Roak BJ, Mason CE, Abbott T, Gupta A, King RA, Pauls DL, Tischfield JA, Heiman GA, Singer HS, Gilbert DL, Hoekstra PJ, Morgan TM, Loring E, Yasuno K, Fernandez T, Sanders S, et al. (2010) L-histidine decarboxylase and Tourette's syndrome. N Engl J Med 362:1901–1908. CrossRef Medline

Fisher RA (1922) On the interpretation of χ2 from contingency tables, and the calculation of P. J R Stat Soc Ser B 85:87–94. CrossRef

Freund CL, Gregory-Evans CY, Furukawa T, Papaioannou M, Looser J, Ploder L, Bellingham J, Ng D, Herbrick JA, Duncan A, Scherer SW, Tsui LC, Loutradis-Anagnostou A, Jacobson SG, Cepko CL, Bhattacharya SS, McInnes RR (1997) Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. Cell 91:543–553. CrossRef Medline

Gal A, Orth U, Baehr W, Schwinger E, Rosenberg T (1994) Heterozygous missense mutation in the rod cGMP phosphodiesterase beta-subunit gene in autosomal-dominant stationary night blindness. Nat Genet 7:551. Medline

Galter D, Westerlund M, Carmine A, Lindqvist E, Sydow O, Olson L (2006) LRRK2 expression linked to dopamine-innervated areas. Ann Neurol 59:714–719. CrossRef Medline

Garbett K, Ebert PJ, Mitchell A, Lintas C, Manzi B, Mirnics K, Persico AM (2008) Immune transcriptome alterations in the temporal cortex of subjects with autism. Neurobiol Dis 30:303–311. CrossRef Medline

Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP, Imielinski M, Frackelton EC, Reichert J, Crawford EL, Munson J, Sleiman PM, Chiavacci R, Annaiah K, Thomas K, Hou C, et al (2009) Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature 459:569–573. CrossRef Medline

Görlich A, Antolin-Fontes B, Ables JL, Frahm S, Slimak MA, Dougherty JD, Ibañez-Tallon I (2013) Reexposure to nicotine during withdrawal increases the pacemaking activity of cholinergic habenular neurons. Proc Natl Acad Sci U S A 110:17077–17082. CrossRef Medline

Han BS, Iacovitti L, Katano T, Hattori N, Seol W, Kim KS (2008) Expression of the LRRK2 gene in the midbrain dopaminergic neurons of the substantia nigra. Neurosci Lett 442:190–194. CrossRef Medline

Honda M, Eriksson KS, Zhang S, Tanaka S, Lin L, Salehi A, Hesla PE, Maehlen J, Gaus SE, Yanagisawa M, Sakurai T, Taheri S, Tsuchiya K, Honda Y, Mignot E (2009) IGFBP3 colocalizes with and regulates hypocretin (orexin). PLoS One 4:e4254. CrossRef Medline

Hupe M, Li MX, Gertow Gillner K, Adams RH, Stenman JM (2013) Evaluation of TRAP-sequencing technology with a versatile conditional mouse model. Nucleic Acids Res. Advance online publication. Retrieved December 18, 2013. doi:10.1093/nar/gkt995. CrossRef Medline

Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A, Kendall J, Grabowska E, Ma B, Marks S, Rodgers L, Stepansky A, Troge J, Andrews P, Bekritsky M, Pradhan K, et al (2012) De novo gene disruptions in children on the autistic spectrum. Neuron 74:285–299. CrossRef Medline

Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, Sousa AM, Pletikos M, Meyer KA, Sedmak G, Guennel T, Shin Y, Johnson MB, Krsnik Z, Mayer S, Fertuzinhos S, Umlauf S, Lisgo SN, Vortmeyer A, Weinberger DR, et al (2011) Spatio-temporal transcriptome of the human brain. Nature 478:483–489. CrossRef Medline

Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, Yokochi M, Mizuno Y, Shimizu N (1998) Mutations in the parkin gene

cause autosomal recessive juvenile parkinsonism. Nature 392:605–608. CrossRef Medline

Knight ZA, Tan K, Birsoy K, Schmidt S, Garrison JL, Wysocki RW, Emiliano A, Ekstrand MI, Friedman JM (2012) Molecular profiling of activated neurons by phosphorylated ribosome capture. Cell 151:1126–1137. CrossRef Medline

Kuhn A, Thu D, Waldvogel HJ, Faull RL, Luthi-Carter R (2011) Population-specific expression analysis (PSEA) reveals molecular changes in diseased brain. Nat Methods 8:945–947. CrossRef Medline

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559. CrossRef Medline

Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segrè AV, Speliotes EK, Wheeler E, Soranzo N, Park JH, Yang J, Gudbjartsson D, et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467:832–838. CrossRef Medline

Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, Chen L, Chen L, Chen TM, Chin MC, Chong J, Crook BE, Czaplinska A, Dang CN, Datta S, Dee NR, et al (2007) Genome-wide atlas of gene expression in the adult mouse brain. Nature 445:168–176. CrossRef Medline

Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21:3448–3449. CrossRef Medline

Marín O (2012) Interneuron dysfunction in psychiatric disorders. Nat Rev Neurosci 13:107–120. CrossRef Medline

Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V, Polak P, Yoon S, Maguire J, Crawford EL, Campbell NG, Geller ET, Valladares O, Schafer C, Liu H, Zhao T, et al. (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 485:242–245. CrossRef Medline

Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH (2008) Functional organization of the transcriptome in human brain. Nat Neurosci 11:1271–1282. CrossRef Medline

O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, Turner EH, Stanaway IB, Vernot B, Malig M, Baker C, Reilly B, Akey JM, Borenstein E, Rieder MJ, Nickerson DA, et al. (2012) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485:246–250. CrossRef Medline

Ozgül RK, Siemiatkowska AM, Yücel D, Myers CA, Collin RW, Zonneveld MN, Beryozkin A, Banin E, Hoyng CB, van den Born LI, Bose R, Shen W, Sharon D, Cremers FPM, Klevering BJ, den Hollander AI, Corbo JC, Consortium ERD (2011) Exome sequencing and cis-regulatory mapping identify mutations in MAK, a gene encoding a regulator of ciliary length, as a cause of retinitis pigmentosa. Am J Hum Genet 89:253–264. CrossRef Medline

Pardal R, López-Barneo J (2012) Neural stem cells and transplantation studies in Parkinson's disease. Adv Exp Med Biol 741:206–216. CrossRef Medline

Pardo CA, Xu Z, Borchelt DR, Price DL, Sisodia SS, Cleveland DW (1995) Superoxide-dismutase is an abundant component in cell-bodies, dendrites, and axons of motor-neurons and in a subset of other neurons. Proc Natl Acad Sci U S A 92:954–958. CrossRef Medline

Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, Horvath S, Geschwind DH (2013) Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell 155:1008–1021. CrossRef Medline

Peyron C, Faraco J, Rogers W, Ripley B, Overeem S, Charnay Y, Nevsimalova S, Aldrich M, Reynolds D, Albin R, Li R, Hungs M, Pedrazzoli M, Padigaru M, Kucherlapati M, Fan J, Maki R, Lammers GJ, Bouras C, Kucherlapati R, et al. (2000) A mutation in a case of early onset narcolepsy and a generalized absence of hypocretin peptides in human narcoleptic brains. Nat Med 6:991–997. CrossRef Medline

Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Almeida J, Bacchelli E, Bader GD, Bailey AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bölte S, Bolton PF, et al. (2010) Functional impact of global rare copy number variation in autism spectrum disorders. Nature 466:368–372. CrossRef Medline

Poewe W, Antonini A, Zijlmans JC, Burkhard PR, Vingerhoets F (2010)

Levodopa in the treatment of Parkinson's disease: an old drug still going strong. Clin Interv Aging 5:229–238. CrossRef Medline

Ramsköld D, Wang ET, Burge CB, Sandberg R (2009) An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput Biol 5:e1000598. CrossRef Medline

Rodriguez-Oroz MC (2010) Deep brain stimulation for advanced Parkinson's disease. Lancet Neurol 9:558–559. CrossRef Medline

Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, et al. (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485:237–241. CrossRef Medline

Saunders JA, Tatard-Leitman VM, Suh J, Billingslea EN, Roberts TP, Siegel SJ (2013) Knockout of NMDA receptors in parvalbumin interneurons recreates autism-like phenotypes. Autism Res 6:69–77. CrossRef Medline

Sgadò P, Genovesi S, Kalinovsky A, Zunino G, Macchi F, Allegra M, Murenu E, Provenzano G, Tripathi PP, Casarosa S, Joyner AL, Bozzi Y (2013) Loss of GABAergic neurons in the hippocampus and cerebral cortex of Engrailed-2 null mutant mice: implications for autism spectrum disorders. Exp Neurol 247:496–505. CrossRef Medline

Shen-Orr SS, Tibshirani R, Khatri P, Bodian DL, Staedtler F, Perry NM, Hastie T, Sarwal MM, Davis MM, Butte AJ (2010) Cell type-specific gene expression differences in complex tissues. Nat Methods 7:287–289. CrossRef Medline

Shi J, Walker MG (2007) Gene set enrichment analysis (GSEA) for interpreting gene expression profiles. Curr Bioinform 2:133–137. CrossRef

Stichel CC, Augustin M, Kühn K, Zhu XR, Engels P, Ullmer C, Lübbert H (2000) Parkin expression in the adult mouse brain. Eur J Neurosci 12:4181–4194. CrossRef Medline

Suzuki K, Sugihara G, Ouchi Y, Nakamura K, Futatsubashi M, Takebayashi K, Yoshihara Y, Omata K, Matsumoto K, Tsuchiya KJ, Iwata Y, Tsujii M, Sugiyama T, Mori N (2013) Microglial activation in young adults with autism spectrum disorder. JAMA Psychiatry 70:49–58. CrossRef Medline

Taheri S, Zeitzer JM, Mignot E (2002) The role of hypocretins (orexins) in sleep regulation and narcolepsy. Annu Rev Neurosci 25:283–313. CrossRef Medline

Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98:5116–5121. CrossRef Medline

Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH (2011) Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 474:380–384. CrossRef Medline

Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, Salyakina D, Imielinski M, Bradfield JP, Sleiman PM, Kim CE, Hou C, Frackelton E, Chiavacci R, Takahashi N, Sakurai T, Rappaport E, Lajonchere CM, Munson J, Estes A, et al. (2009) Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature 459:528–533. CrossRef Medline

Wills S, Rossi CC, Bennett J, Martinez Cerdeño V, Ashwood P, Amaral DG, Van de Water J (2011) Further characterization of autoantibodies to GABAergic neurons in the central nervous system produced by a subset of children with autism. Mol Autism 2:5. CrossRef Medline

Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, Reilly SK, Lin L, Fertuzinhos S, Miller JA, Murtha MT, Bichsel C, Niu W, Cotney J, Ercan-Sencicek AG, Gockley J, Gupta AR, Han W, He X, Hoffman EJ, et al (2013) Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell 155:997–1007. CrossRef Medline

Xu X, Nehorai A, Dougherty JD. (2013) Cell type specific analysis of human brain transcriptome data to predict alterations in cellular composition. Syst Biomed 1:1–10. CrossRef

Zeng H, Shen EH, Hohmann JG, Oh SW, Bernard A, Royall JJ, Glattfelder KJ, Sunkin SM, Morris JA, Guillozet-Bongaarts AL, Smith KA, Ebbert AJ, Swanson B, Kuan L, Page DT, Overly CC, Lein ES, Hawrylycz MJ, Hof PR, Hyde TM, et al. (2012) Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. Cell 149:483–496. CrossRef Medline

Zhou P, Zhang Y, Ma Q, Gu F, Day DS, He A, Zhou B, Li J, Stevens SM, Romo D, Pu WT (2013) Interrogating translational efficiency and lineage-specific transcriptomes using ribosome affinity purification. Proc Natl Acad Sci U S A 110:15395–15400. CrossRef Medline